Incremental and Adaptive Feature Exploration over Time Series Stream

Jingwei ZUO, Karine ZEITOUNI, Yehia TAHER

2nd Juin, 2022

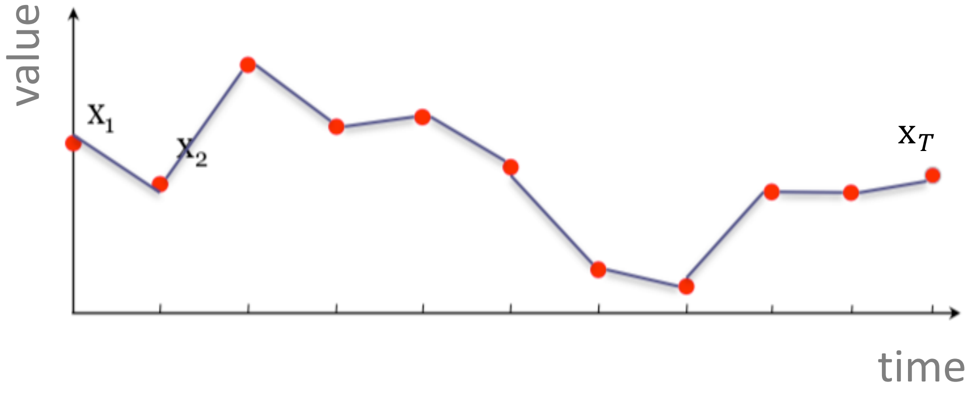
Seminar with Miners team, LIMOS, UCA

Context & definitions

- Time series
  - Sequence of points ordered by time

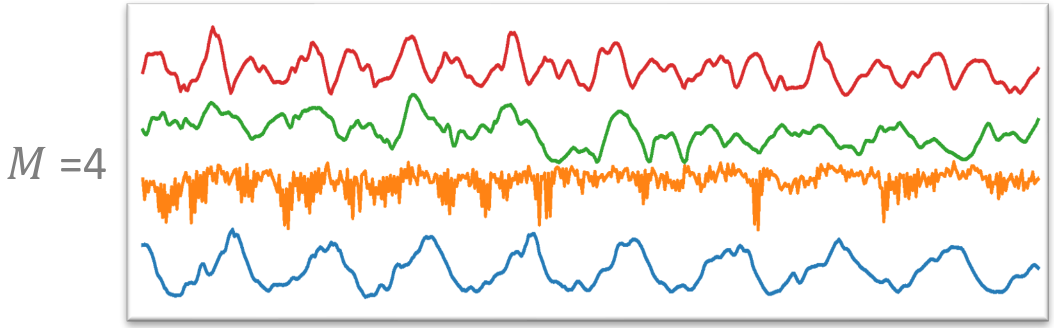
Univariate Time Series (UTS)

$$x_1, \dots, x_T \in \mathcal{R}^M, M = 1$$



Multivariate Time Series (MTS)

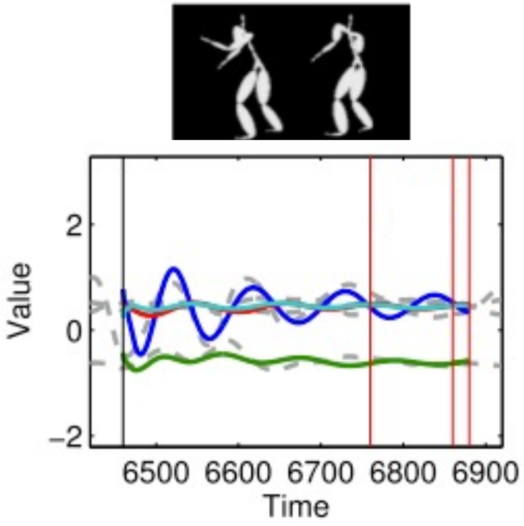
$$x_1, \dots, x_T \in \mathcal{R}^M, M > 1$$



Data from SHL-Huawei dataset

Context & definitions

- Streaming context:
  - real-valued data flow (e.g., real-time sensor data)



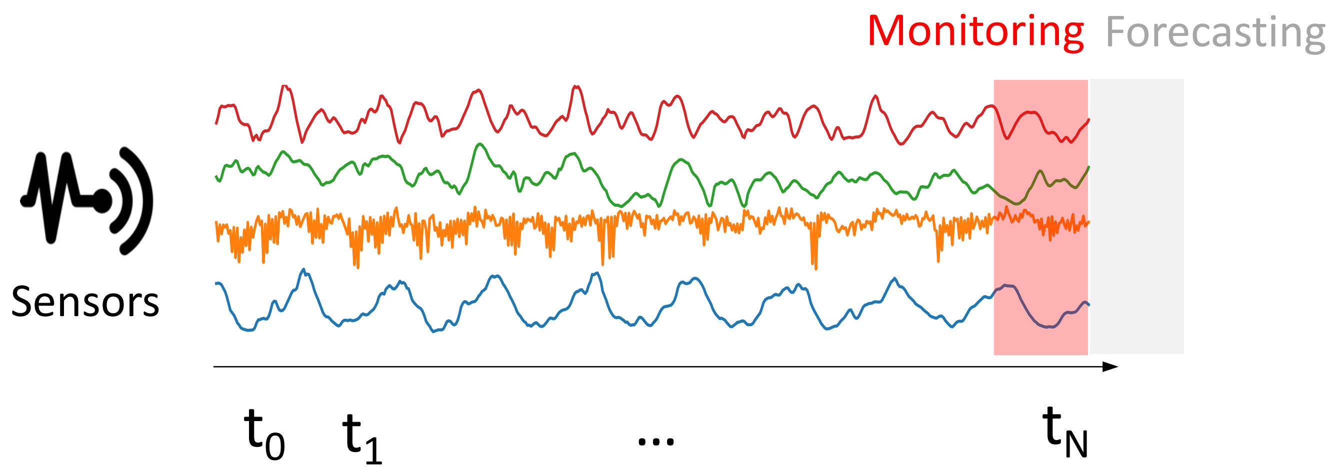
Real-time sensor data

- Time series in streaming context
  - Historical time series, i.e., offline time series
  - Streaming time series
  - Time series stream

Context & definitions

- Streaming time series
  - A continuous input data stream where each instance is a real-valued data: S=($t_1, t_2, \ldots, t_N$), where N is the time of the most recent input value.

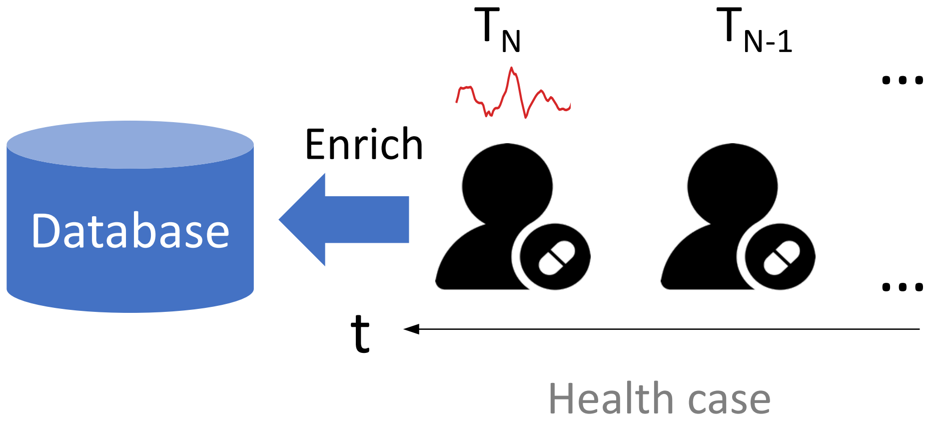
- Use cases:
  - Online monitoring
  - Real-time forecasting



Context & definitions

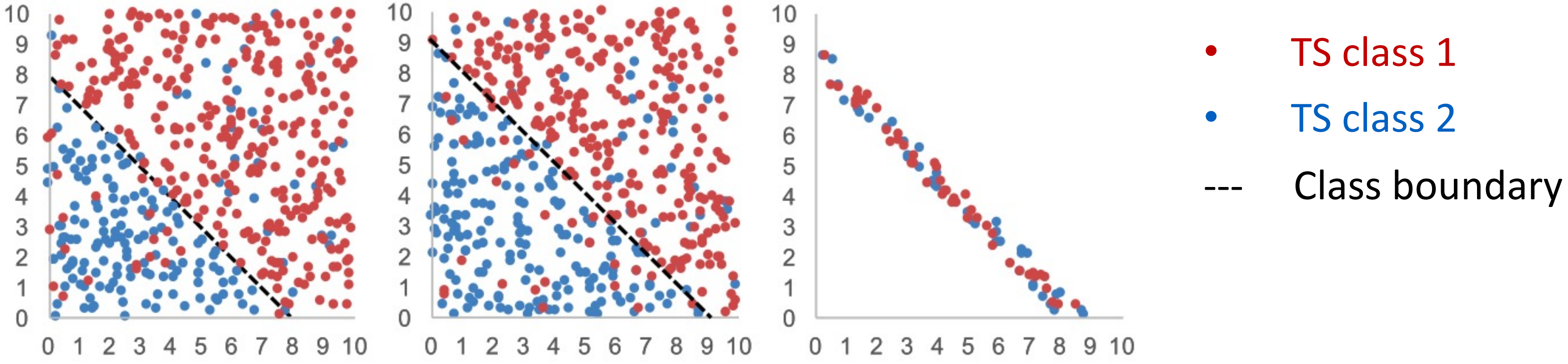
- Time series stream (our context)
  - A continuous input data stream where each instance is a time series: $S_{TS} = (T_1, T_2, \ldots, T_N)$, notice that N increases with each new time-tick.

- Use cases:
  - Medical domain (e.g., ECG)
  - Astronomy discovery (e.g., Star Light Curves)



Problem statement

- Complex temporal relationships in time series stream
  - Infinite length
  - Feature evolution
  - Concept drift



Data at time t　　Data at time t+1　　Data in drift region

Objectives

- TS features in streaming context
  - **Interpretability**: visually interpretable
  - **Incrementality**: feature extraction is incremental with new-coming instances [Feature Evolution]
  - **Adaptability**: adaptive to the evolving data distribution [Concept Drift]

- Learning model
  - **Scalability**

- Mainly designed for Time Series Classification (TSC) Task
  - **Training online**, classification on-line or off-line

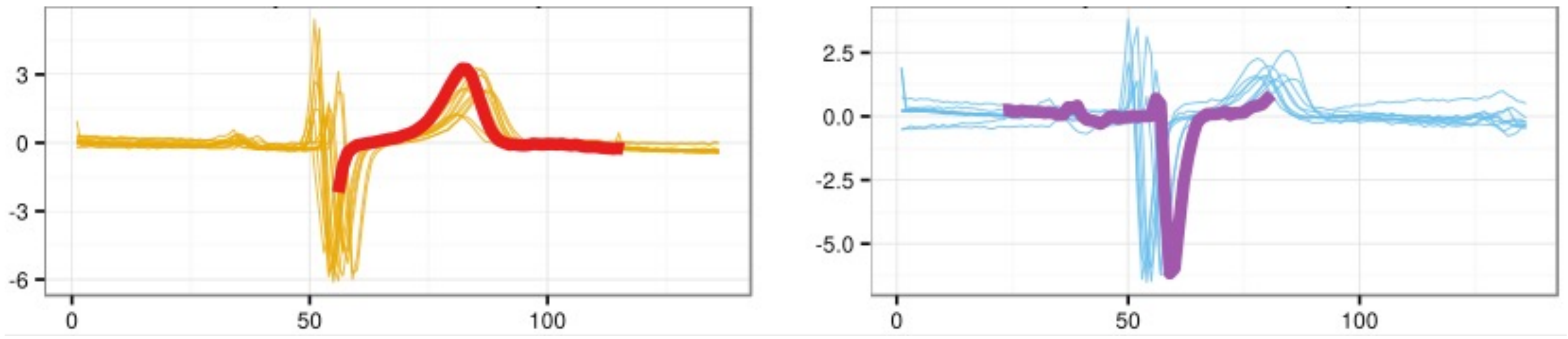
Related work

- Time series representation for classification

Feature representations	**Classifier example**	**Related work**
Raw representation	1-NN	**1NN-ED,** **1NN-DTW** and its variants
Statistic summary	SVM or tree-based	**TSF** [Deng et al., Inf. Sci. 2013]
Deep representations	Neural Networks	**mWDN** [Wang et al., KDD’18], **InceptionTime** [Fawaz et al., DMKD’19], **LSTM-FCN** [Farim et al., arXiv’19]
Feature/model ensembles	Ensemble classifier	**BOSS** [Schäfer, DMKD’15] and its variants, **HIVE-COTE** [Lines et al., ICDM’17], **TDE** [Middlehurst et al., PKDD’20]
Local patterns	SVM or tree-based	**RPM** [Wang and Lin, EDBT’16], **Shapelet** [Ye and Keogh, KDD’09] and its variants

Why Shapelet[1] in our context?

- Definition
  - A representative shape in time series which is capable of distinguishing one class from the others

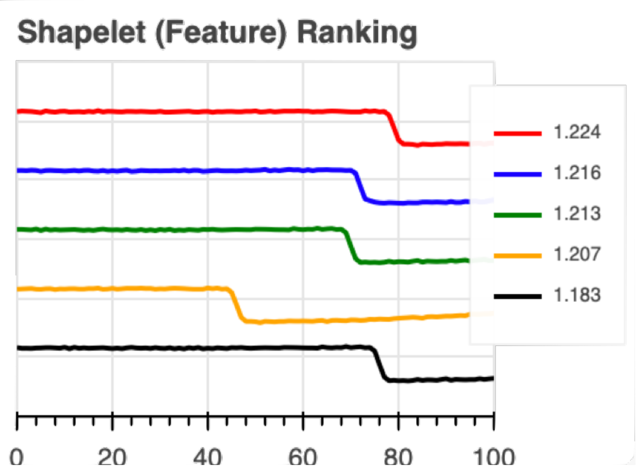


Most representative Shapelets in two classes from ECGFiveDays
[Wang and Lin, EDBT'16]

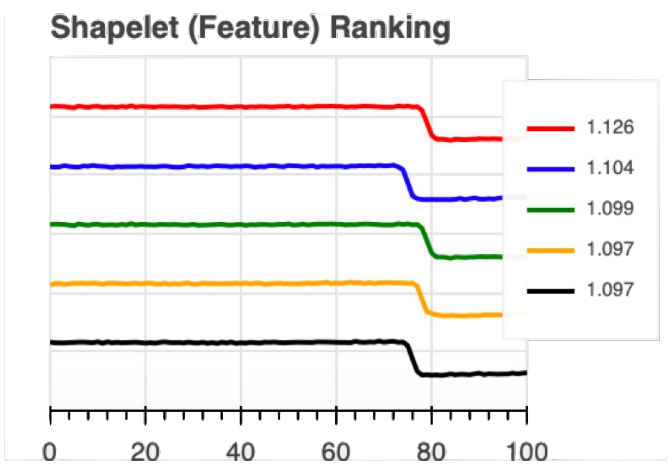
1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009

Why Shapelet[1] in our context?

- Explainable for Feature Evolution in time series stream



T_1 = 20
Dataset *Trace*[2] (class 2)



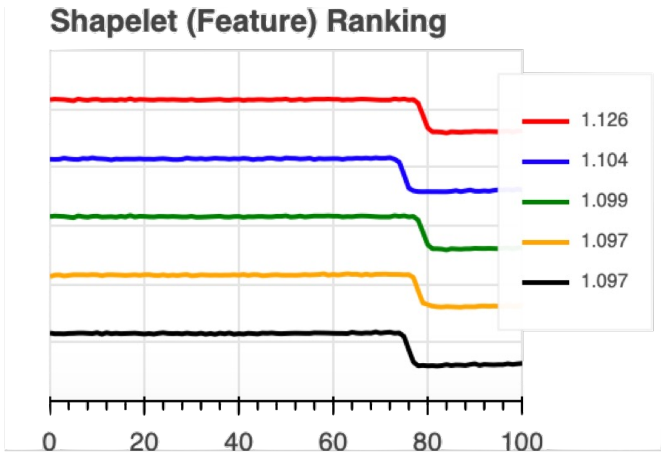
T_2 = 100
Dataset *Trace*[2] (class 2)

Feature Evolution
between t $\in [20, 100]$

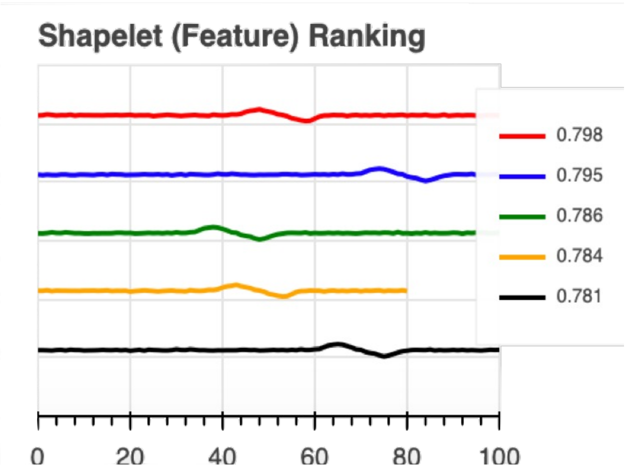
1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009
2. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

Why Shapelet[1] in our context?

- Explainable for Concept Drift in time series stream



T_2 = 100
Dataset *Trace*[2] (class 2)



T_3 = 200
Dataset *Trace*[2] (class 2)

Concept Drift between
$t \in [100, 200]$

1. L. Ye and E. Keogh. “Time series shapelets: A New Primitive for Data Mining.” In Proc. SIGKDD 2009
2. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

Shapelet-based methods

Shapelet-based methods

Extraction-based

- Highly interpretable (decision-tree)

Figure from [Ye and Keogh, KDD’09]

Learning-based

- End-to-end (gradient-based learning)
- Generally not interpretable

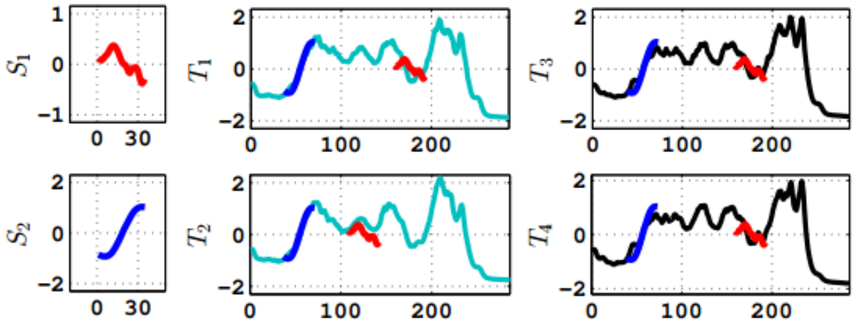


Figure from [Grabocka et al., KDD’14]

Algorithm for Shapelet Extraction

- Distance Profile & Matrix Profile[1]

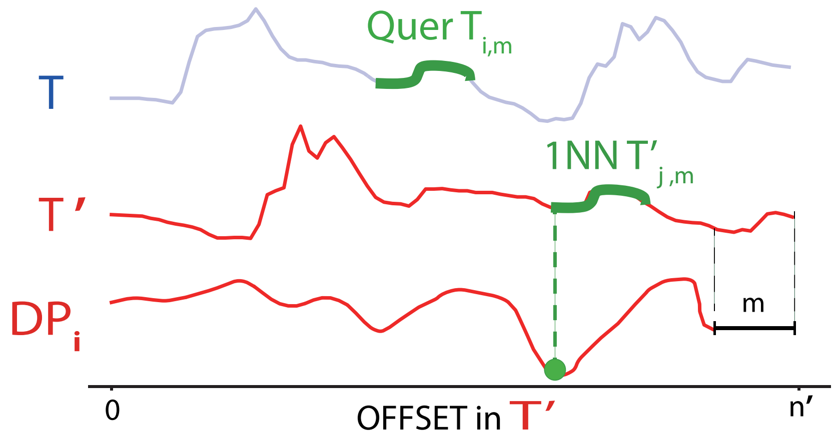


Figure 2.1: ***Distance Profile*** *between Query $T_{i,m}$ and target time series T', where n' is the length of T'. $DP_{i,j}$ can be considered as a meta TS annotating target T'*

➢ Find the Nearest Neighbor of the Query

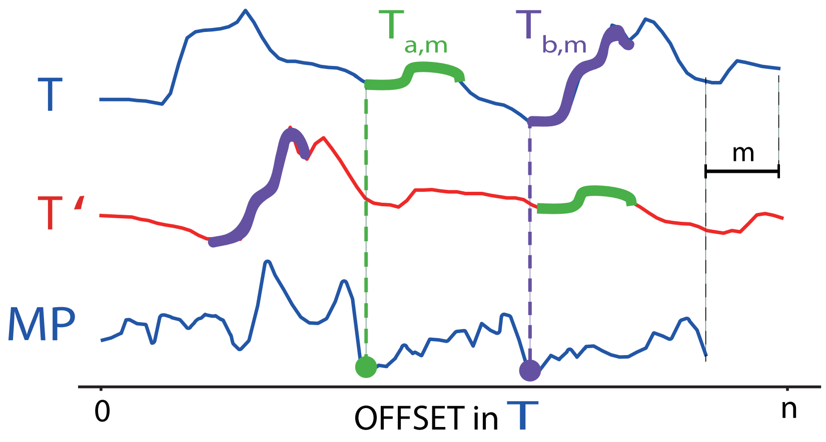


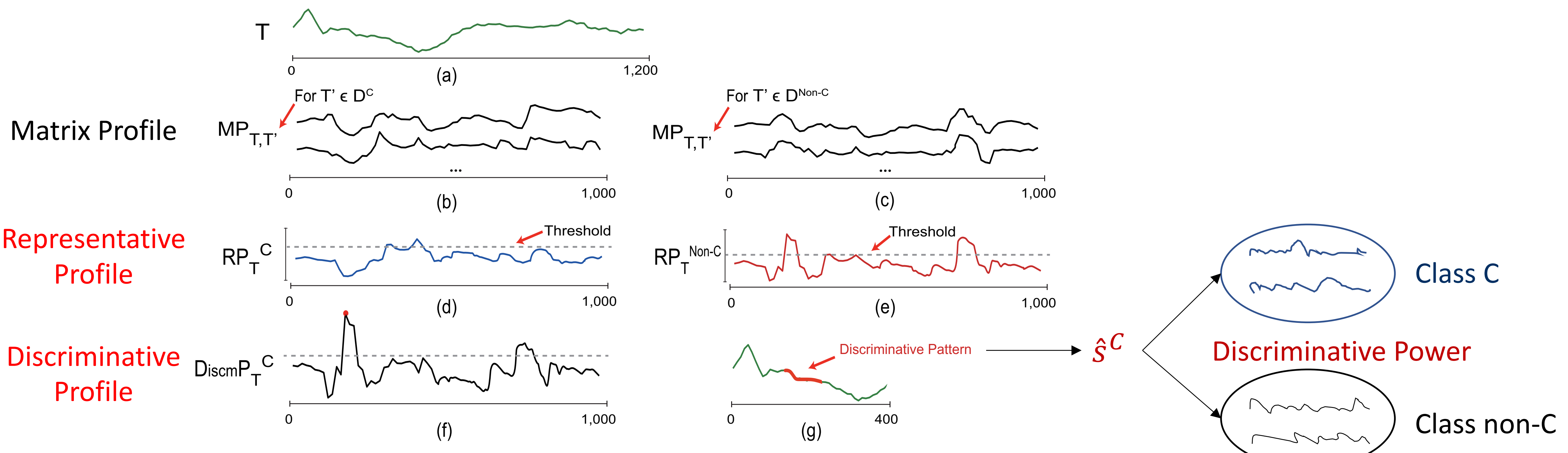
Figure 2.2: ***Matrix Profile*** *between Source T and Target T', where n is the length of T. Intuitively, MP_i shares the same offset as source T*

➢ Find the closest pairs between two TS

1. Chin-Chia Michael Yeh et al. "Matrix Pro le I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets." In Proc. ICDM 2016

Proposal - SMAP

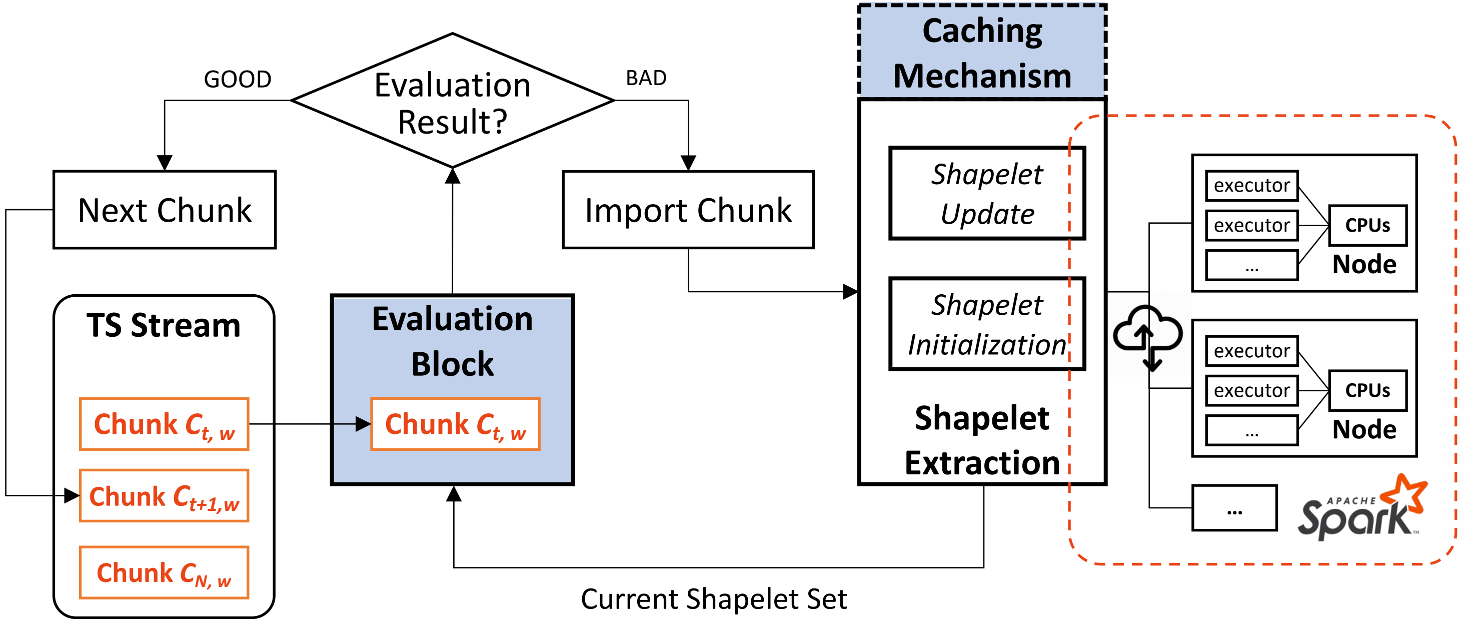
- SMAP[1] : Shapelet Extraction on Matrix Profile



1. J. Zuo, K. Zeitouni and Y. Taher, Exploring interpretable features for large time series with SE4TeC. In Proc. EDBT 2019

Proposal - Incremental version of SMAP

- ISMAP[1] : **Incremental** and **adaptive** Shapelet Extraction on Matrix Profile



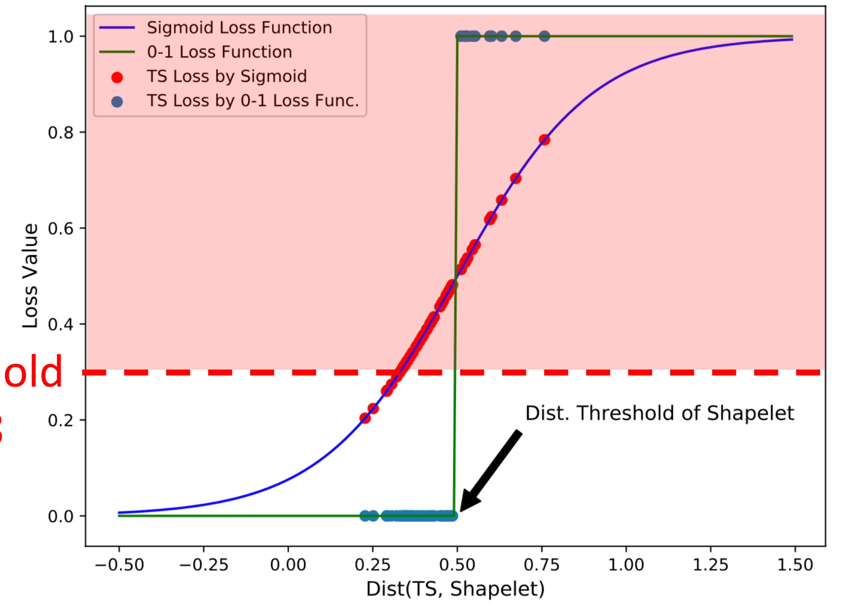
Test-then-Train strategy

1. J. Zuo, K. Zeitouni and Y. Taher, Incremental and Adaptive Feature Exploration over Time Series Stream, IEEE Big Data 2019

ISMAP - Evaluation Block

Shapelet Evaluation

e.g.,
loss threshold
$\Delta = 0.3$



Shapelet Evaluation over newly input TS instances

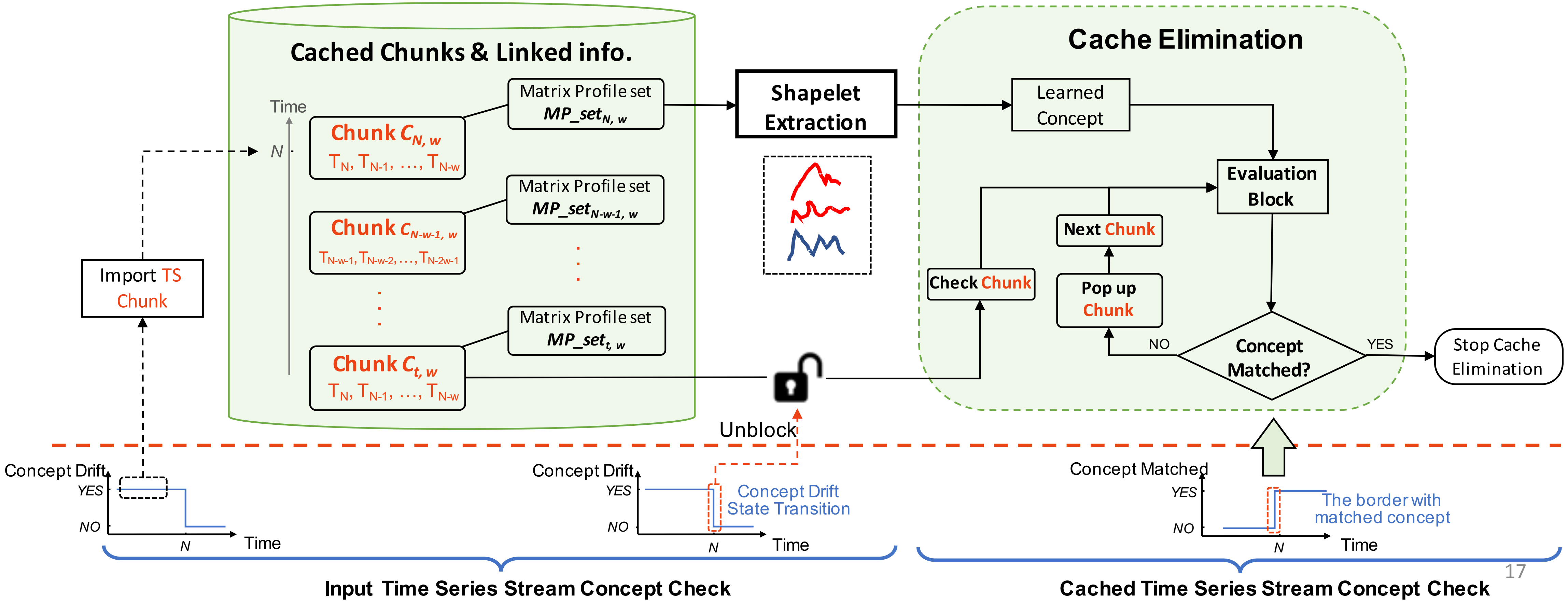
Concept Drift detection

Consider the evaluation loss as a signal

- Page-Hinkey (PH) Test: initially designed for change point detection in signal processing.
- λ: *PH threshold* to detect a Concept Drift
- $Concept\ Drift = \begin{cases} True, & PH_N \geq \lambda \\ False, & otherwise \end{cases}$

ISMAP - Elastic Caching Mechanism

- One-pass algorithm
  - Only conserve the data under the current concept to be learned
- Conserve the historical Shapelets in the out-of-date concepts (optional)



Experiments

Research Questions:

- RQ1. Incremental learning with ISMAP
  - Stable-concept time series stream
  - To validate the incremental behavior

→

Datasets:

- 14 datasets from UCR Archive[1]

Baselines (Shapelet Tree classifiers):

- Information Gain (IG) [Ye and Keogh, KDD'09]
- Kruskall-Wallis (KW), Mood's Median (MM) [Lines and Bagnall, IDEAL'12]

- RQ2. Adaptive learning with ISMAP
  - Drifting-concept time series stream
  - To validate the drift detection behavior and elastic caching mechanism

→

Datasets:

- Synthetic *Trace* and *ECG5000* datasets[1]:
  - Randomly put noise for Data Augmentation
  - Two concept drifts are inserted in each dataset

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

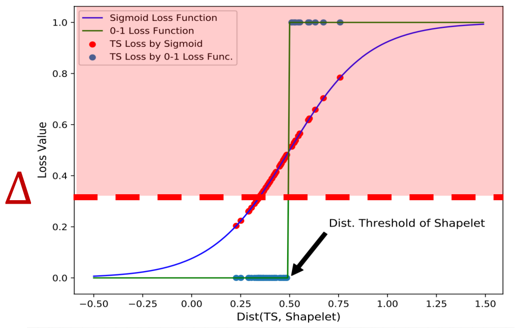
RQ1. Incremental learning with ISMAP

- Incremental behavior
  - Captured by $Compression\ Ratio = \frac{N_{import}}{N_{training}}$
- Possible to combine with other TS classifiers:
  - Shapelet Transform [Lines et al., KDD'12]
  - HIVE-COTE [Lines et al., ICDM'16]

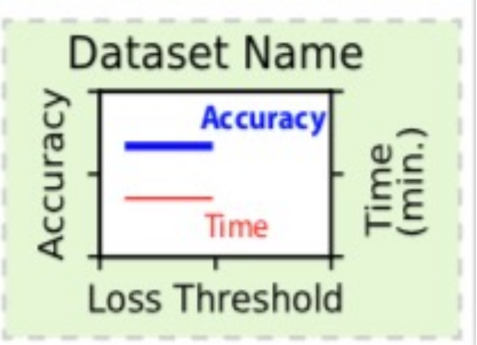
Type	Name	Train/Test	Class	Length	IG	KW	MM	ISMAP(best)	Para. (Δ)	Comp. Ratio
Simulated	SyntheticControl	300/300	6	60	**0.9433**	0.9000	0.8133	0.7007	0.35	46.7%
Sensor	Trace	100/100	4	275	0.9800	0.9400	0.9200	**1**	0.5, 0.45	26.0%
	MoteStrain	20/1252	2	84	0.8251	0.8395	0.8395	**0.9169**	0.45	60.0%
	SonyAIBO.I	20/601	2	70	0.8453	0.7281	0.7521	**0.9151**	0.4	95.0%
	SonyAIBO.II	27/953	2	65	0.8457	-	-	**0.8583**	0.4	63.0%
	ItalyPower.	67/1029	2	24	0.8921	0.9096	0.8678	**0.9466**	0.45	25.4%
ECG	ECG5000	500/4500	5	140	0.7852	-	-	**0.9109**	0.4	9.4%
	ECGFiveDays	23/861	2	136	0.7747	0.8721	0.8432	**0.9826**	0.4	51.2%
	TwoLeadECG	23/1189	2	82	0.8507	0.7538	7657	**0.9337**	0.5	47.8%
Images	Symbols	25/995	6	398	0.7799	0.5568	0.5799	**0.8113**	0.35	96.0%
	Coffee	28/28	2	286	**0.9643**	0.8571	0.8671	0.9286	0.4	78.6%
	FaceFour	24/88	4	350	0.8409	0.4432	0.4205	**0.9886**	except 0.45	62.5%
	DiatomSize.	16/306	4	345	0.7222	0.6111	0.4608	**0.8758**	0.5	50.0%
Motion	GunPoint	50/150	2	150	0.8933	0.9400	0.9000	**0.9733**	0.45	42.0%

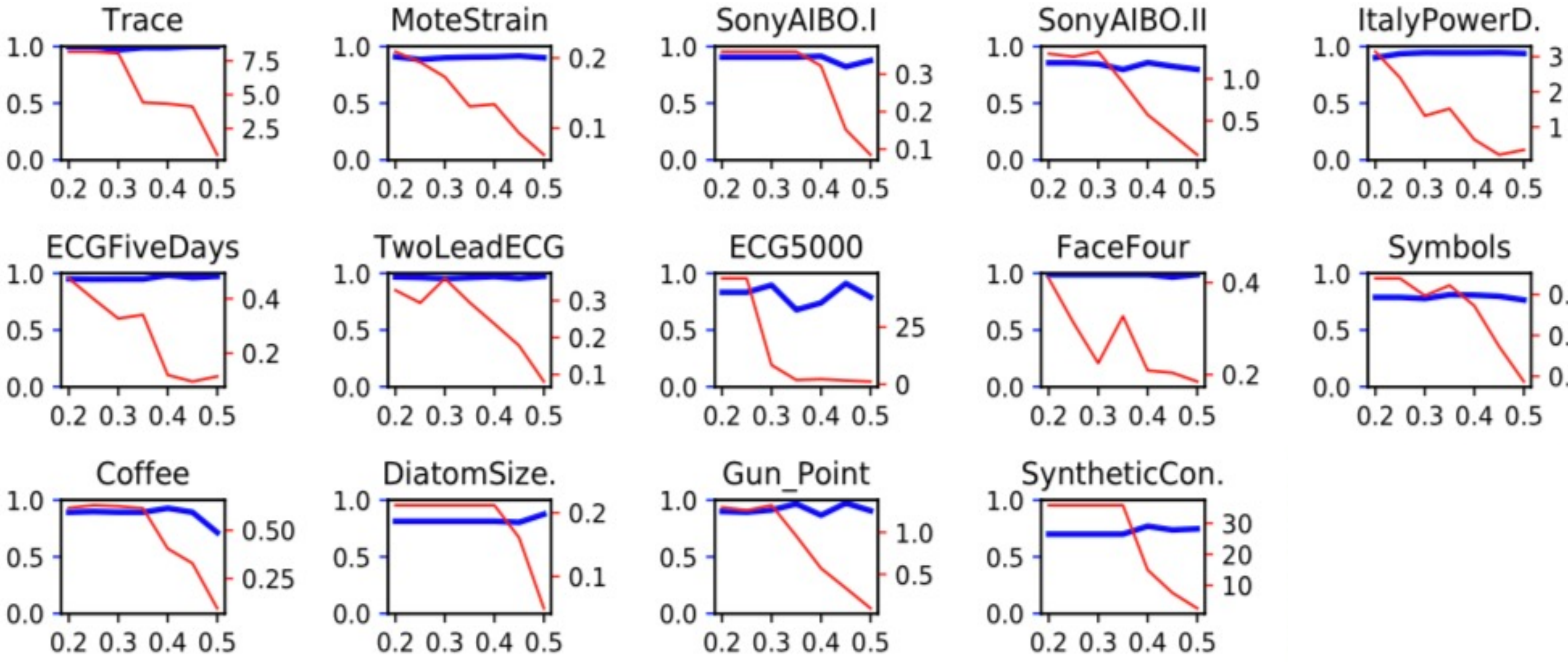
1. J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in Proc. SIGKDD 2012
2. J. Lines, S. Taylor, and A. Bagnall, "Hive-cote: The hierarchical vote collective of transformation-based ensembles for time series classification," IEEE ICDM 2016

RQ1. Incremental learning with ISMAP



- Trade-off between Accuracy and Loss Threshold Δ





In theory

- Loss threshold ↗, the efficiency ↗, the accuracy ↘

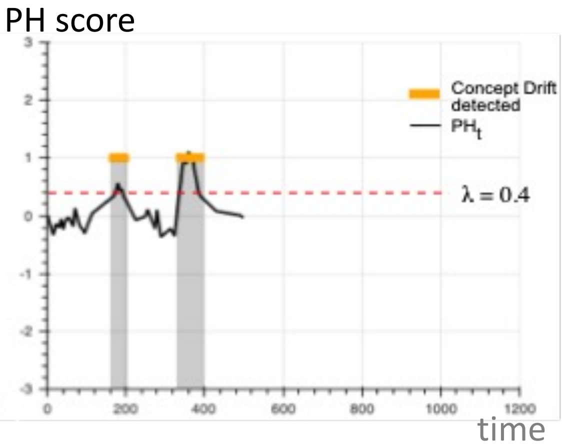
In practice

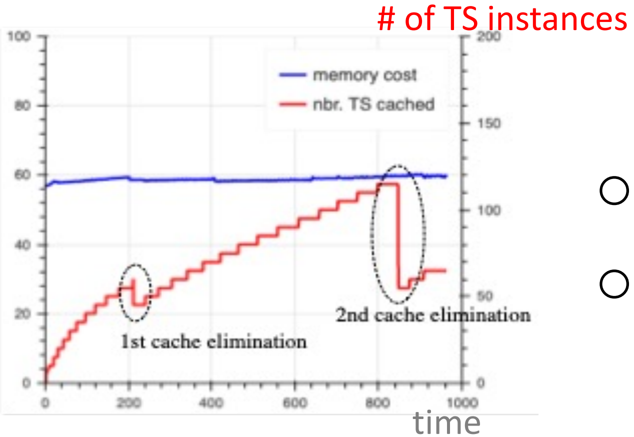
- The highest accuracy falls in the range Δ ∈ [0.35,0.45].

RQ2. Adaptive learning with ISMAP

- Concept drift detection & Elastic caching mechanism[1]

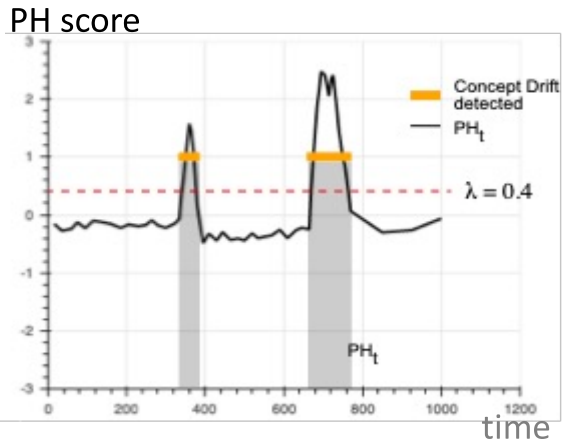
ECG5000

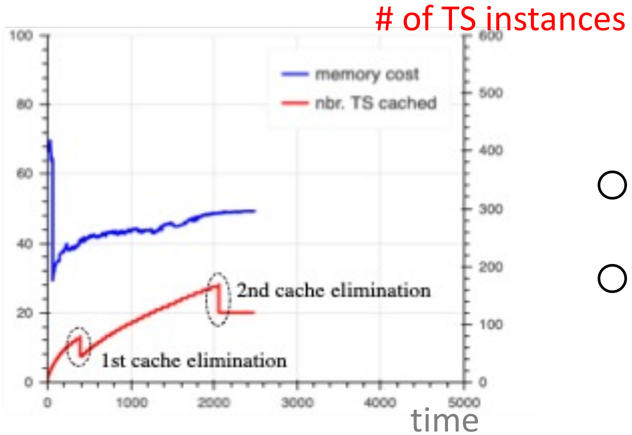




- Two concept drifts detected
- 65 of 500 instances cached

Trace





- Two concept drifts detected
- 120 of 1000 instances cached

1. J. Zuo, K. Zeitouni, and Y. Taher, "ISETS: Incremental Shapelet Extraction from Time Series Stream", demo paper in ECML-PKDD'19

ISMAP - Conclusion

- Shapelet representation is natively interpretable for explaining the feature evolution and concept drift in the time series stream.
- Our proposal *ISMAP* extracts incremental and adaptive Shapelets from the time series stream
- Our proposed *elastic caching mechanism* handles the infinite time series stream.
- ISMAP is applicable in the scenarios where:
  - New TS instances enrich the learned concept
  - New TS instances may lead to Concept Drift

Github page